European Journal of Cancer 123 (2019) 116–117

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejcancer.com

Editorial

Overestimated treatment effects in randomised phase II trials: What's up doctor?

Phase II clinical trials of experimental treatments play an essential role in drug development. Historically, trials were conducted sequentially, starting from a phase I trial in which dose and safety are evaluated advancing to phase II, looking for some sign of efficacy on a short-term end-point such as tumour shrinkage to screen out inefficacious experimental agents, and to large-scale randomised phase III trials evaluating properly the efficacy of a new treatment on objective clinical end-points. To expedite this process, phase II trials historically used single-arm designs to treat patients with experimental therapies only [1]. However, the overall success rate of drug development programs in oncology from the start of phase I to registration has been estimated recently to be only 3.4%, much lower compared with that of other diseases [2]. Part of this low success rate has been argued to be due to suboptimal use of randomised trial designs in the phase II space.

Liang *et al.* [3] performed a literature search of randomised phase II studies published in twelve major journals in the period 2006–2015. They tracked down subsequent phase III studies of the same experimental treatment in the same setting and identified a total of 57 phase II–phase III trial pairs. Among the 57 phase III studies, 11 studies have not (yet) been published in the medical literature, and the results were obtained at ClinicalTrial.gov or from meeting abstracts. Compared with the matched phase III studies, treatments effects of progression-free survival (PFS) were on average 26% larger in the phase II studies (ratio of hazard ratio [rHR] = 0.74, 95% confidence interval [CI]: 0.68–0.80), and similarly, treatment effects of overall survival (OS) were on average 27% larger (rHR = 0.73, 95% CI: 0.66–0.79). Unsurprisingly, in phase II trials, treatment effects based on subgroup analysis were even more overestimated. In univariable screening, the strongest predictor of phase III trial success was to have a positive phase II trial, which confirms a previous analysis in a large database of anticancer drugs [4]. The authors consider some explanations for this, including publication bias of phase III randomised controlled trials, multiple phase II trials being conducted for a drug and greater heterogeneity in patient selection for phase III trials.

We welcome this interesting and thorough piece of work that raises important issues for consideration. A major reason for the overestimation of treatment effects in randomised phase II trials on time-to-event end-points such as PFS and OS is purely statistical: the existence of a phase III trial implies that a promising phase II trial was performed, which leads to optimism bias in phase II trials for which a phase III trial followed. To illustrate this concept, we performed a small simulation study. First, we randomly sampled a true treatment effect from a historical (lognormal) distribution of treatment effects of drugs in oncology derived from a meta-analysis of randomised clinical trials led by the United States National Cancer Institute for more than 50 years [5]. We generated a randomised phase II trial (of a 70 patients, the median size in the study by Liang *et al.* [3]) using exponential distributions, a median survival time of 1 year in the control arm and the sampled true treatment effect. When a 'significant' treatment effect was identified using a log-rank test, at a one-sided 2.5% level, we also generated a matched randomised phase III trial (of a size of 608). This process was repeated until 57 pairs were identified to replicate the analysis strategy of Liang *et al.* [3]. On average, among those generated phase II trials that were significant at the 2.5% level, treatment effects were 12% larger in the randomised phase II studies than in the phase III counterparts (rHR = 0.88, 95% CI: 0.84–0.93). One could also draw a parallel with treatment effects estimated

DOI of original article: https://doi.org/10.1016/j.ejca.2019.08.006.

https://doi.org/10.1016/j.ejca.2019.09.023
0959-8049/© 2019 Elsevier Ltd. All rights reserved.

in randomised phase III trials that stopped early for large treatment benefits during interim analysis. A systematic review showed that compared with consecutive meta-analyses or large untruncated phase III studies, treatment effect sizes in phase III trials that stopped early for benefit were on average 29% larger (rHR = 0.71, 95% CI: 0.65−0.77) [6], which is remarkably similar to the rHR estimated in the present study and confirms that early results on a small number of patients included in a randomised trial can be randomly high.

Other possible contributors for the overestimation of treatment effects are manifold. A well-known issue is publication bias, i.e. 'positive' phase II studies with favourable results are more likely to be published than others. There is also the potential issue of p-hacking in the analysis of a phase II trial. If one plays the devil's advocate, one could assume a little bit more flexibility in statistical modelling in a randomised phase II study as compared to a large-scale phase III study which strictly controls the type 1 error. This is likely to be even more in the case of subgroup analyses, where it is common to report analyses without making it clear they are exploratory. Finally, as the authors acknowledge, there can also be differences between phase II and matched phase III studies through the use of different doses, schedules or combination of therapies or more heterogeneity in trial participants or implementation of interventions in the large-scale phase III study. Particular statistical methods have even been proposed to properly discount the treatment effect estimate from a phase II trial in the design of the phase III one [7].

In conclusion, we concur with the view of Liang *et al.* [3] that caution must be taken in interpreting treatment effects estimated in published medium-sized randomised phase II trials and especially those obtained beyond the primary prespecified analyses. We would further emphasise that all of these issues are even more likely to be present when the phase II trial is a single-arm one. Grayling *et al.* [8] found that 28.2% of published randomised phase II oncology trials were positive compared with 72.7% of non-randomised ones. Randomisation in phase II settings is an essential building block of innovative clinical trial designs that are needed to enhance the probability of success of drug development programs such as in multiarm multistage trials and biomarker-based clinical trials (umbrella, basket, and so on) [9].

Conflict of interest statement

The authors declare no related conflict of interest.

References

[1] Jung S. Current issues in phase II cancer clinical trials. In: Halabi S, Michiels S, editors. Textbook of clinical trials in oncology a statistical perspective. New York: Chapman and Hall/CRC; 2019.
[2] Wong CH, Siah KW, Lo AW. Estimation of clinical trial success rates and related parameters. Biostatistics 2018;20:273−86.
[3] Liang F, Wu Z, Mo M, Zhou C, Shen J, Wang Z, et al. Comparison of treatment effect from randomized controlled phase II trials and subsequent phase III trials using identical regimens in the same treatment setting. Eur J Cancer 2019;121:19−28. https://doi.org/10.1016/j.ejca.2019.08.006.
[4] Jardim DL, Groves ES, Breitfeld PP, Kurzrock R. Factors associated with failure of oncology drugs in late-stage clinical development: a systematic review. Cancer Treat Rev 2017;52:12−21.
[5] Bayar MA, Le Teuff G, Michiels S, Sargent DJ, Le Deley MC. New insights into the evaluation of randomized controlled trials for rare diseases over a long-term research horizon: a simulation study. Stat Med 2016;35:3245−58.
[6] Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q, et al. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. J Am Med Assoc 2010;303:1180−7.
[7] Kirby S, Burke J, Chuang-Stein C, Sin C. Discounting phase 2 results when planning phase 3 clinical trials. Pharm Stat 2012;11:373−85.
[8] Grayling MJ, Dimairo M, Mander AP, Jaki TF. A review of perspectives on the use of randomization in phase II oncology trials. J Natl Cancer Inst 2019. https://doi.org/10.1093/jnci/djz126. pii: djz126.
[9] Renfro LA, Sargent DJ. Statistical controversies in clinical research: basket trials, umbrella trials, and other master protocols: a review and examples. Ann Oncol 2016;28:34−43.

Stefan Michiels*
Service de Biostatistique et D'Épidémiologie, Gustave Roussy, Université Paris-Saclay, 94805 Villejuif, France

CESP INSERM U1018, Université Paris-Sud, Université Paris-Saclay, 94805 Villejuif, France

James Wason
Institute of Health and Society, Newcastle University, Newcastle Upon Tyne, UK

MRC Biostatistics Unit, University of Cambridge, Cambridge, UK

**Corresponding author:* Service de Biostatistique et d'Épidémiologie, Gustave Roussy, Université Paris-Saclay, 94805 Villejuif, France.
E-mail address: stefan.michiels@gustaveroussy.fr

20 September 2019
Available online 1 November 2019